Nonsyndromic cleft palate

Ishorst, Nina; Francheschelli, Paola; Böhmer, Anne C.; Khan, Mohammad Faisal J.; Heilmann-Heimbach, Stefanie; Fricker, Nadine

Published in:
Birth Defects Research

DOI:
10.1002/bdr2.1213

Publication date:
2018

Document Version
Peer reviewed version

Nonsyndromic cleft palate: An association study at GWAS candidate loci in a multi-ethnic sample

Journal:	*Birth Defects Research*
Manuscript ID	BDR-17-0234.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	19-Jan-2018
Complete List of Authors:	Ishorst, Nina; University of Bonn, Institute of Human Genetics; Life and Brain Center, Department of Genomics Francheschelli, Paola; University of Ferrara, Department of Biomedical and Specialty Surgical Sciences, Section of Medical Biochemistry, Molecular Biology and Genetics Böhmer, Anne; University of Bonn, Institute of Human Genetics; Department of Genomics, Life and Brain Center, University of Bonn Khan, Mohammad Faisal J.; University of Ferrara, Department of Biomedical and Specialty Surgical Sciences, Section of Medical Biochemistry, Molecular Biology and Genetics Heilmann-Heimbach, Stefanie; University of Bonn, Institute of Human Genetics; Department of Genomics, Life and Brain Center, University of Bonn Fricker, Nadine; University of Bonn, Institute of Human Genetics; Department of Genomics, Life and Brain Center, University of Bonn Little, Julian; Univ Ottawa, Epidemiol & Commun Hlth Steegers-Theunissen, Régine; Erasmus MC, University Medical Center, Obstetrics and Gynecology Peterlin, Borut; University Medical Center Ljubljana , Clinical Institute of Medical Genetics, Department of Obstetrics & Gynecology Nowak, Stefanie; University of Bonn, Institute of Human Genetics Martini, Markus; Rheinische Friedrich Wilhelms Universitat Bonn, Department of Oral and Maxillo-Facial-Plastic Surgery Kruse, Teresa; University of Cologne, Department of Orthodontics Dunsche, Anton; Clinics Karlsruhe, Department of Oral and Maxillo-Facial Surgery Kreusch, Thomas; Asklepios Klinik Nord, Heidberg, Department of Oral and Maxillofacial Surgery, Head and Neck Centre Gölz, Lina; University of Bonn, Department of Orthodontics Aldhorae, Khalid Ahmed; Thamar University, Orthodontic Department, College of Dentistry Halboub, Esam; Jazan University, College of Dentistry, Devision of Oral Medicine and Diagnostic Sciences, Department of Maxillofacial Surgery and Diagnostic Sciences Reutter, Heiko; University of Bonn, Institute of Human Genetics; University of Bonn, Neonatology Mossey, Peter; University of Dundee, Dental School Nöthen, Markus; University of Bonn, Institute of Human Genetics; Life and Brain GmbH, Department of Genomics Rubini, Michele; University of Ferrara, Department of Biomedical and Specialty Surgical Sciences Ludwig, Kerstin; Institute of Human Genetics, Department of Genomics

	Knapp, Michael; University of Bonn, Institute of Medical Biometry, Informatics, and Epidemiology Mangold, Elisabeth; University of Bonn, Institute of Human Genetics
Key Words:	congenital malformation, common variants, imputed genome-wide association study, nonsyndromic cleft palate only, candidate loci, association study

Nonsyndromic cleft palate: An association study at GWAS candidate loci in a multi-ethnic sample

Authors:

Nina Ishorst [1,2]*, Paola Francheschelli [3]*, Anne C. Böhmer [1,2], Mohammad Faisal J. Khan [3], Stefanie Heilmann-Heimbach [1,2], Nadine Fricker [1,2], Julian Little [4], Regine P.M. Steegers-Theunissen [5], Borut Peterlin[6], Stefanie Nowak [2], Markus Martini [7], Teresa Kruse [8], Anton Dunsche [9], Thomas Kreusch [10], Lina Gölz [11], Khalid Aldhorae [12], Esam Halboub [13], Heiko Reutter [2,14], Peter Mossey [15], Markus M. Nöthen [1,2], Michele Rubini [3], Kerstin U. Ludwig [1,2], Michael Knapp [16], Elisabeth Mangold [2]

Affiliations:

[1] Department of Genomics, Life&Brain Center, University of Bonn, Bonn, Germany

[2] Institute of Human Genetics, University of Bonn, Bonn, Germany

[3] Department of Biomedical and Specialty Surgical Sciences, Section of Medical Biochemistry, Molecular Biology and Genetics, University of Ferrara, Ferrara, Italy

[4] School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa, Ottawa, Ontario, Canada

[5] Department of Obstetrics and Gynaecology, Department of Pediatrics, Division Neonatology Erasmus MC, University Medical Center, Rotterdam, The Netherlands

[6] Clinical Institute of Medical Genetics, Department of Obstetrics & Gynecology, University Medical Center Ljubljana, Ljubljana, Slovenia

[7] Department of Oral and Maxillo-Facial-Plastic Surgery, University of Bonn, Bonn, Germany

[8] Department of Orthodontics, University of Cologne, Cologne, Germany

[9] Clinics Karlsruhe, Department of Oral and Maxillo-Facial Surgery, Karlsruhe, Germany

[10] Department of Oral and Maxillofacial Surgery, Head and Neck Centre, Asklepios Klinik Nord-Heidberg, Hamburg , Germany

[11] Department of Orthodontics, University of Bonn, Bonn, Germany

[12] Orthodontic Department, College of Dentistry, Thamar University, Thamar, Yemen

[13] Devision of Oral Medicine and Diagnostic Sciences, Department of Maxillofacial Surgery and Diagnostic Sciences, College of Dentistry, Jazan University, Jazan, Saudi Arabia

[14] Department of Neonatology, Children's Hospital, University of Bonn, Bonn, Germany

[15] Dental Hospital, University of Dundee, Dundee, UK

[16] Institute of Medical Biometry, Informatics and Epidemiology, University of Bonn, Bonn, Germany

* contributed equally

Grant information: The study was supported by the German Research Foundation (grant FOR 423

and individual grants MA 2546/3-1, KR 1912/7-1, NO 246/6-1, and WI 1555/5-1, LU-1944/3-1), and by the European Science Foundation (grant 09-RNP-023 (EMRC), ESF Network for Orofacial Clefts Research, Prevention and Treatment: EUROCleftNet)

Running title: Candidate loci association study in nsCPO

Correspondence to:

Dr. Elisabeth Mangold, Institute of Human Genetics, University of Bonn, Sigmund-Freud-Straße 25, 53127 Bonn, Germany. Phone: +49 (0)228 287 51008; Fax: +49 (0) 228 287 51011 Email: e.mangold@uni-bonn.de

Abstract (250 words):

Background: Nonsyndromic cleft palate only (nsCPO) is a common and multifactorial form of orofacial clefting. In contrast to successes achieved for the other common form of orofacial clefting, i.e., nonsyndromic cleft lip with/without cleft palate (nsCL/P), genome wide association studies (GWAS) of nsCPO have identified only one genome wide significant locus. Aim of the present study was to investigate whether common variants contribute to nsCPO and, if so, to identify novel risk loci.

Methods: We genotyped 33 SNPs at 27 candidate loci from 2 previously published nsCPO GWAS in an independent multi-ethnic sample. It included: (i) a family-based sample of European ancestry (n=212); and (ii) two case/control samples of Central European (n=94/339) and Arabian ancestry (n=38/231), respectively. A separate association analysis was performed for each genotyped dataset, and meta-analyses were performed.

Results: After association analysis and meta-analyses, none of the 33 SNPs showed genome-wide significance. Two variants showed nominally significant association in the imputed GWAS dataset and exhibited a further decrease in P-value in a European and an overall meta-analysis including imputed GWAS data, respectively (rs395572: $P_{MetaEU}=3.16x10^{-4}$; rs6809420: $P_{MetaAll}=2.80x10^{-4}$).

Conclusion: Our findings suggest that there is a limited contribution of common variants to nsCPO. However, the individual effect sizes might be too small for detection of further associations in the present sample sizes. Rare variants may play a more substantial role in nsCPO than in nsCL/P, for which GWAS of smaller sample sizes have identified genome-wide significant loci. Whole-exome/genome sequencing studies of nsCPO are now warranted.

Keywords: congenital malformation, nonsyndromic cleft palate only, common variants, imputed genome-wide association study, candidate loci, association study

Introduction:

Orofacial clefting is one of the most common forms of congenital malformation, and displays substantial phenotypic variation. The two most common forms of orofacial clefting are cleft lip with/without cleft palate (CL/P) and cleft palate only (CPO) (Mangold et al., 2011). These two types are distinct in terms of both epidemiological findings and developmental origin, and can occur in a syndromic or nonsyndromic form. In the latter, no additional physical or cognitive disabilities are present (Mangold et al., 2011). In Europe, the estimated prevalence of true nonsyndromic clefts is 1 in 1,000 for nonsyndromic CL/P (nsCL/P) and 1 in 2,400 for nonsyndromic CPO (nsCPO), suggesting that nsCL/P is the more common form of nonsyndromic clefting in this population (Mangold et al., 2011). Research suggests that nsCL/P and nsCPO are multifactorial disorders, with both genetic and environmental factors contributing to their development. Heritability estimates of approximately 90% have been reported for both phenotypes (Grosen et al., 2012).

To date, research has identified a total of 39 genome-wide significant risk loci for nsCL/P across different populations. Of these, 37 loci were identified via genome-wide association studies (GWAS), meta-analyses of GWAS data, or follow-up studies (Rahimov et al., 2008; Birnbaum et al., 2009; Moreno et al., 2009; Beaty et al., 2010; Mangold et al., 2010; Ludwig et al., 2012; Beaty et al., 2013; Sun et al., 2015; Leslie et al., 2016b; Ludwig et al., 2016; Yu et al., 2017; Leslie et al., 2017; Ludwig et al., 2017).

In contrast, only two independent GWAS (Beaty et al., 2011; Leslie et al., 2016a), and one meta-analysis of both studies (Leslie et al., 2017), have been published for nsCPO. One of these studies (Leslie et al., 2016a) identified the first - and as yet only - genome-wide significant risk locus for nsCPO at chromosome 1p36, a finding that was then independently detected in a sequencing study (Mangold et al., 2016). Notably, this missense variant in the *grainyhead-like 3* gene (*GRHL3*) showed no association with nsCL/P (Mangold et al., 2016).

The issue of whether nsCPO and nsCL/P share a common genetic etiology is a matter of ongoing debate. However, available epidemiological- and molecular data suggest that the shared genetic etiology of these two frequent forms of clefting is limited. Ludwig et al. (2017) demonstrated that with the exception of the *FOXE1* locus (9q22), none of the nsCL/P risk loci known at the time of the analysis exhibited a significant association in an imputed nsCPO GWAS dataset (Ludwig et al., 2017). The finding for *FOXE1* was unsurprising, since previous research has established that this locus is associated with both traits (Moreno et al., 2009; Ludwig et al., 2014; Leslie et al., 2017). A polygenic score approach in the Ludwig et al. (2017) study also showed that nsCL/P loci in sum do not

contribute to nsCPO etiology (Ludwig et al., 2017). These findings are consistent with the results of a current study of Moreno-Uribe et al. (2017).

Aim of the present study was to investigate whether common variants contribute to nsCPO and, if so, to identify novel risk loci. To that end we investigated genetic association of candidate loci derived from the already mentioned imputed nsCPO GWAS dataset (Ludwig et al., 2017). Additionally, we included candidate single nucleotide polymorphisms (SNPs) described in another recent imputed nsCPO GWAS (Leslie et al., 2016a). Association analyses were performed in a European trio sample, and two independent case/control samples from Central Europe and Yemen, respectively. Additionally, we investigated a possible overlap in signals between nsCL/P in nsCPO by checking 15 novel nsCL/P risk loci (Yu et al., 2017; Leslie et al., 2017) for association in the imputed nsCPO dataset (Ludwig et al., 2017). With this investigation we completed the analysis performed by Ludwig et al. (2017) for the 24 risk loci known at that time.

Methods:

Imputed GWAS Dataset (Ludwig et al., 2017)

The imputed nsCPO GWAS data that were followed up in the present study were based on a GWAS that comprised approximately 498,000 genotypes of each member of 550 nuclear nsCPO trios (Beaty et al., 2011). Ludwig et al. (2017) obtained dbGaP data of this nsCPO GWAS and performed genome-wide imputation for 446 trios of European (n=212) and Asian (n=234) ancestry using IMPUTE2 on the basis of 1000 genomes haplotypes (phase 3) (The 1000 Genomes Project Consortium 2015). Statistical analyses were performed for around 8.38 million variants with an info-score of >0.4 and a minor allele frequency (MAF) of > 1% using the FBAT dosage approach (Cobat et al., 2014). In addition to the main imputed dataset of genotypes from 446 trios ($EU+Asia_{Beaty}$), a sub-dataset was generated comprising the genotypes of 212 European trios only (EU_{Beaty}).

Sample

The genotyped sample comprised two independent case/control samples and one trio sample recruited in the context of different studies. Further information on gender distribution and origin is given in Table S1.

The first case/control sample comprised 94 nsCPO index patients and 339 population-matched controls ($EU_{Mangold}$). The cases were drawn from a large sample of nonsyndromic cleft families of Central European ancestry (recruitment method described in Mangold et al., 2010). The population-matched controls were volunteer blood donors also of Central European ancestry (Mangold et al., 2010).

The second case/control sample ($Arab_{Aldhorae}$) was recruited within the context of surgical outreach programs, as described elsewhere (Aldhorae et al., 2014). The $Arab_{Aldhorae}$ sample comprised 38 nsCPO cases and 231 blood donors of Arab ethnicity from Yemen.

The family-based sample comprised 212 nuclear trios with an nsCPO index patient (EU_{Mossey}). These families were recruited from different areas of Europe (Great Britain, Italy, The Netherlands, Slovenia, Slovakia, Hungary, Estonia, and Bulgaria) in the context of the EUROCRAN/ITALCLEFT study (Mossey et al., 2017). Nonsyndromic status of affecteds was confirmed clinically at the point of recruitment to the various studies.

The present study was approved by the ethics committees of the participating institutions, and informed consent was obtained from all participants, or if not adult by parents or legal guardians, prior to inclusion.

Genotyping

DNA from the $EU_{Mangold}$ sample (patients and controls) had been isolated using standard salting out procedures (Mangold et al., 2010). Isolation of DNA from $Arab_{Aldhorae}$ blood samples (patients and controls) had been performed using the Chemagic Magnetic Seperation Module I (Perkin Elmer chemagen Technologies GmbH, Baesweiler, Germany), in accordance with the manufacturer´s instructions (Aldhorae et al., 2014). Genomic DNA from the EU_{Mossey} sample had been extracted from peripheral venous blood samples using the QIAamp Blood DNA Mini kit (QIAGEN, Germantown, MD, US) (Mossey et al., 2017).

DNA was quantified using NanoDrop (PeqLab Thermo Scientific, Wilmington, DE, US). Genotyping of selected SNPs was performed with the multiplex MassARRAY system (Agena Bioscience, Hamburg, Germany), as based on end-point PCR coupled with MALDI-ToF. Two plexes were designed using the Assay Design Suite v2.0, and data analysis was performed using the Typer Analysis software (both: Agena, Hamburg, Germany). Primers were synthesized by Metabion (Planegg, Germany). Primer sequences are available upon request. One SNP was genotyped in both plexes as an internal control.

Selection of SNPs for Genotyping

SNPs with P-values of $< 1x10^{-5}$ in at least one of the GWAS datasets ($EU+Asia_{Beaty}$, EU_{Beaty}) were selected from the imputed GWAS data (Table S2). Of these, SNPs with an info-score < 0.8 were excluded from further analysis.

For loci in which more than 1 SNP fulfilled the selection criteria we selected suitable backup SNPs. SNPs that failed assay design were replaced by the second best hit, in terms of p-value and info-score, at the respective locus. If only one SNP fulfilling the criteria was available at a given locus and failed assay design, a backup SNP in high linkage disequilibrium (LD) ($r^2 > 0.99$) was selected from the

imputed dataset. In this special case also variants with a P-value slightly above the set threshold of $1x10^{-5}$ were considered. The selection procedure resulted in 25 SNPs at 19 loci (Table S2; Table S3). Initially, 13 candidate SNPs (P-value $<1x10^{-5}$) derived from Leslie et al. (2016a) were to be forwarded for genotyping in the present three independent samples, without considering their association in the imputed datasets (EU_{Beaty} and $EU+Asia_{Beaty}$). However, only eight SNPs at eight loci could be included into the genotyping assay (Table S4). Thus, a total of 33 SNPs at 27 loci were selected for genotyping (Table S5).

Statistical Analysis

For the case/control samples, the standard chi-square test was used to test for deviation from Hardy-Weinberg Equilibrium (HWE). The threshold for significant deviation from HWE was set at $P < 1x10^{-4}$ in controls and $P < 1x10^{-6}$ in cases.

Using the two-sided Armitage trend test, each SNP was tested for a difference in genotype distribution between cases and controls. In the EU_{Mossey} trios, single-marker association analysis was performed using the transmission disequilibrium test (TDT), as implemented in the FAMHAP software package (Becker and Knapp, 2004).

Data evaluation comprised: (i) a separate analysis of each genotyped sample ($EU_{Mangold}$, $Arab_{Aldhorae}$, EU_{Mossey}); (ii) a z-score based meta-analysis and a fixed-effects meta-analysis of all genotyped samples (Willer et al., 2010) ($Meta_{geno}$); (iii) a z-score based meta-analysis (Willer et al., 2010) of all European samples ($EU_{Mangold}$, EU_{Mossey}, EU_{Beaty}; named $Meta_{EU}$); and (iv) z-score based meta-analysis (Willer et al., 2010) of all samples ($EU_{Mangold}$, $Arab_{Aldhorae}$, EU_{Mossey}, $EU+Asia_{Beaty}$; termed $Meta_{All}$).

Relative risks were calculated for each of the genotyped samples and their meta-analysis based on the fixed effects approach (Table S6). EU_{Beaty} and $EU+Asia_{Beaty}$ (Ludwig et al., 2017) were imputed using the FBAT dosage approach (Cobat et al., 2014). This method does not allow for calculation of relative risks, thus this data is not available for $Meta_{EU}$ and $Meta_{All}$.

A workflow of the present study is shown in Figure 1.

P-values of < 0.05 were considered nominally significant.

P-values given as $P_{corrected}$ were Bonferroni corrected for multiple testing. In the $EU_{Mangold}$ and EU_{Mossey} data the number of 33 genotyped SNPs was considered for Bonferroni correction (Table S5). In the $Arab_{Aldhorae}$ dataset the number of 32 SNPs was considered (Table S5). For the analysis of 15 nsCL/P loci top SNPs from Yu et al. (2017) and Leslie et al. (2017) in the imputed nsCPO dataset we considered the number of 15 markers for Bonferroni correction (Table S7). P-values of < 0.05 were considered significant after correction for multiple testing.

LD values were determined using the web-based SNAP tool (Broad Institute).

Quality Control (QC) in the Genotyping Step

For the internal control SNP used in both plexes, genotypes were compared and checked for agreement. Duplicates were identified by comparing individuals with shared genotypes, and removed from the genotyping dataset. In addition, samples with >2 missing genotypes (corresponding to an average call rate of 94%) and/or >2 Mendelian inconsistencies in the trio sample, were excluded.

Analysis of nsCL/P loci top SNPs and associated regions of Yu et al. (2017) and Leslie et al. (2017) in EU_{Beaty} and $EU+Asia_{Beaty}$

To complete the analysis of potential signal-overlap for nsCL/P and nsCPO performed in Ludwig et al. (2017), the imputed nsCPO dataset was inspected on both a locus-wide and lead SNP level for the novel loci identified in 2017 by Yu et al. (n= 14) and Leslie et al. (n= 1).

Results:

Imputation resulted in 8,384,634 and 8,384,636 markers (info-score ≥ 0.4; MAF > 1%) in EU_{Beaty} and $EU+Asia_{Beaty}$, respectively (Ludwig et al., 2017). No genome-wide significant association for nsCPO was reported (Ludwig et al., 2017). However, 83 SNPs at 26 loci showed P-values of $< 10^{-5}$ in at least one analysis. The following had info-scores of <0.8 and were thus excluded: (i) 6 SNPs at 6 loci with only one associated SNP; and (ii) one SNP at a locus with multiple SNPs. This resulted in 76 SNPs at 20 loci. Of these 20 loci, one locus could not be included in the genotyping step due to technical issues. For the remaining 19 loci, the selection procedure resulted in 25 SNPs for genotyping (Table S2). The risk alleles of SNPs selected from the imputed data were identical in both datasets (EU_{Beaty} and $EU+Asia_{Beaty}$). None of the eight SNPs at eight loci selected from Leslie et al. (2016a) showed nominally significant association in either of the imputed datasets: The lowest P-values were obtained for rs3740617 in EU_{Beaty} (P = 0.417) and rs595533 in $EU+Asia_{Beaty}$ (P = 0.087) (Table S4). Genotyping was successful for all 33 markers (representing 27 different loci). In the course of QC, several samples were excluded due to missing genotype calls for more than two SNPs. These comprised the samples of: 7 individuals from $EU_{Mangold}$ (2 patients and 5 controls); 2 individuals from $Arab_{Aldhorae}$ (1 patient, 1 control); and 34 individuals from 27 EU_{Mossey} trios. Furthermore, the analyses identified six EU_{Mossey} trios with more than two Mendelian inconsistencies, and two duplicates (1 control from $Arab_{Aldhorae}$ and 1 EU_{Mossey} trio). These samples were excluded from further analysis. One SNP (rs76706433) was monomorphic in the Yemeni population ($Arab_{Aldhorae}$). This variant was therefore excluded from the analyses of $Arab_{Aldhorae}$, resulting in a set of 32 SNPs for this particular sample. None of the analyzed SNPs showed a deviation from HWE at the selected thresholds.

Genotyped sample and $Meta_{geno}$

Separate analysis of each genotyped sample ($EU_{Mangold}$, $Arab_{Aldhorae}$, EU_{Mossey}) (Figure 1) revealed four SNPs (rs6747560, rs6809420, rs6805813, rs395572) with nominally significant P-values ($P < 0.05$). These comprised one SNP in the $EU_{Mangold}$ sample and three SNPs in the $Arab_{Aldhorae}$ sample (Table 1). At three of these loci, the risk allele was not identical to the risk allele of the imputed GWAS dataset. In the $Arab_{Aldhorae}$ sample, rs6809420 was the only SNP to show both nominally significant association and a consistent effect direction ($P = 0.024$). This result did not withstand correction for multiple testing of the 32 SNPs ($P_{corrected} = 0.752$). The fixed effects-based analysis revealed a relative risk (RR) of 2.41(95% confidence interval (CI), 1.12-5.20) for this SNP (Table S6).

The z-score-based meta-analysis of all three genotyped samples ($Meta_{geno}$) revealed two nominally significant SNPs (rs6747560, rs1283107). Both SNPs showed the opposite effect direction to that observed in the initial datasets (EU_{Beaty} and $EU+Asia_{Beaty}$) (Table 1, Table S8). Notably, in $Meta_{geno}$ rs6809420 showed a tendency towards a nominally significant association ($P_{Metageno} = 0.066$) and the fixed-effects-based meta-analysis revealed an RR of 1.24 (95% CI, 0.97-1.60) (Table S6).

$Meta_{EU}$

In the meta-analysis of all samples with European ancestry ($EU_{Mangold}$, EU_{Mossey}, EU_{Beaty}) (Figure 1), a decrease in the association P-value in comparison to EU_{Beaty} alone was observed for the SNPs rs395572, rs7732608, and rs3740617 (Table 2). Of these, only rs395572 at 3q29 showed a nominally significant association ($P_{EUBeaty} = 0.0017$; $P_{MetaEU} = 3.16 \times 10^{-4}$) (Figure 2). None of the SNPs from Leslie et al. (2016a) showed decreased association P-values in the meta-analysis of all European samples.

$Meta_{All}$

In the z-score-based meta-analysis of all trios from the imputed $EU+Asia_{Beaty}$ dataset and all genotyped samples ($EU_{Mangold}$, $Arab_{Aldhorae}$, EU_{Mossey}) (Figure 1), lower P-values compared to $EU+Asia_{Beaty}$ alone were observed for the SNPs rs6809420, rs7732608, rs9347594, rs3740617, and rs595533 (Table 3). Of these, only two SNPs showed nominally significant association. Rs6809420 was nominally significant in the initial imputed GWAS data ($EU+Asia_{Beaty}$), and the association signal got stronger after combination with the genotyped samples ($P_{EU+AsiaBeaty} = 0.0011$, $P_{MetaAll} = 2.80 \times 10^{-4}$) (Figure 3). Here, $Arab_{Aldhorae}$ (Table 3) made the main contribution in terms of lowest P-value from the genotyped samples as already reported in section 'Genotyped Sample and $Meta_{geno}$'.

The association P-value for rs595533 from Leslie et al. (2016a) became nominally significant following data combination ($P_{EU+AsiaBeaty} = 0.087$, $P_{MetaAll} = 0.0412$) (Table 3).

In comparison to $Meta_{EU}$, the association P-value of rs395577 showed no further decrease in $Meta_{All}$. This is probably due to an inconsistent effect direction for this SNP in the Yemen sample (Table S9, Table S10).

Analysis of nsCL/P loci top SNPs and associated regions of Yu et al. (2017) and Leslie et al. (2017) in EU_{Beaty} and $EU+Asia_{Beaty}$

Analysis of these 15 novel loci revealed no genome wide significant association at either the whole locus or lead SNP level approach. In the locus-wide approach, the lowest P values were obtained for rs73288895 and rs8009664 from Yu et al. (2017) in EU_{Beaty} and $EU+Asia_{Beaty}$, respectively ($P_{EUBeaty}$ = $2.87x10^{-4}$ and $P_{EU+AsiaBeaty}$ = $3.88x10^{-5}$) (Table S11; Table S12). Both variants are located at 14q22.1. According to a SNAP proxy search (Johnson et al., 2008), neither variant is in LD with the lead SNP rs7148069 at the 14q22.1 locus (Yu et al., 2017) in European or Asian reference populations. In the lead SNP approach, rs705704 at 12q13.2 showed a nominally significant P-value in $EU+Asia_{Beaty}$ ($P_{EU+AsiaBeaty}$ = 0.010) (Table S7). However, this did not withstand Bonferroni correction for multiple testing of the 15 tested lead SNPs ($P_{corrected}$ = 0.155).

Discussion:

To date, research has identified only one genome-wide significant nsCPO risk variant, namely rs41268753 at the *GRHL3* locus (1p36) (Leslie et al., 2016a; Mangold et al., 2016). The aim of the present study was to further elucidate the genetic background of nsCPO, in particular whether common variants do contribute to the phenotype.

We tried to detect such common nsCPO risk loci by combining imputed nsCPO GWAS data (Ludwig et al., 2017) with genotyping data from independent samples in two different ways:

In the first approach data of all individuals under study were combined, regardless of ethnicity. This strategy aimed at compensating for small sample sizes and is expected to identify variants that contribute to nsCPO risk in diverse populations, albeit with low effect sizes. This led to the identification of a lead variant at 3q13 (rs6809420) with a frequency of 81.9% in the general population (Non-Finnish Europeans; gnomAD) (Exome Aggregation Consortium, 2016), which is located in an intergenic region. The genes *TUSC1* and *IGSF11* are located 1 Mb upstream and 1.2 Mb downstream of rs6809420, respectively, which suggests that this variant may have a regulatory function (Figure 3). Further support for the association with rs6809420 is provided by the variants rs9829549 ($P_{EU+AsiaBeaty}$ = 0.0013; r^2 = 1) and rs4602427 ($P_{EU+AsiaBeaty}$ = 7.50×10^{-4}, r^2 = 0.94) which lie in perfect/high LD with the lead variant (red dots, Figure 3).

Evidence for a potential regulatory function of genetic regions can be provided by publicly available databases like the Roadmap Epigenomics Consortium (2015). These data show that rs6809420 is located in an enhancer mark (H3K4me1) of CD56+ ectoderm cultured cells derived from human embryonic stem cells, although it should be noted that these cells are derived from ectoderm, and are not identical to the palate-forming neural crest cells. Nevertheless, due to the early embryological time-point of facial development, access to appropriate human tissue samples for the epigenetic or functional follow-up of non-coding regions is problematic (Thieme and Ludwig, 2017). Furthermore, ENCODE transcription factor binding experiments have predicted that rs6809420 leads to the alteration of a p300 binding site (Kheradpour and Kellis, 2014). Together with other translational coactivators, p300 has been shown to exhibit distinct expression patterns during murine craniofacial development (Bhattacherjee et al., 2009). An altered p300 binding site may thus lead to the dysregulation of genes essential to this critical developmental process. Through an interaction with the effects of other variants and/or environmental factors, this may lead to the development of a cleft palate. Altogether, rs6809420 at 3q13 is an interesting finding. Another variant identified by this multi-ethnic approach was rs595533 at 18q22.1. This variant was reported by Leslie et al. (2016a). Only a limited decrease in the P-value was observed following data combination. We do not consider rs595533 an interesting suggestive finding of the present study.

The second approach was done in view of the predominantly European nature of the samples under study. In this 'European only' approach a meta-analysis of all European samples ($Meta_{EU}$) was performed. This ethnicity-specific meta-analysis identified a second suggestive variant (rs395572) with a frequency of 67.4% in the general population (Non-Finnish Europeans; gnomAD) (Exome Aggregation Consortium, 2016). Rs395572is located upstream of the microRNA gene *MIR4797* and of the gene *DLG1* (Figure 2). Recent studies in humans and animal models showed that microRNAs play a role in embryonic development including craniofacial development and might contribute to the risk of orofacial clefting when disrupted in their function (Wang et al., 2013; Ding et al., 2016; Li et al., 2016). Rs395572 may alter expression of *MIR4797*. Alternatively, rs395572 might be in LD with a true causal variant located within the microRNA gene, altering binding properties of miR-4797. *DLG1* encodes the *human discs large protein 1*, which belongs to the family of molecular scaffolding proteins, and is implicated in biological processes such as the regulation of epithelial cell polarity and cell migration (Roberts et al., 2012; Marziali et al., 2015). Moreover, a study in mice demonstrated that murine *Dlg1* is required for craniofacial development, and leads to craniofacial abnormalities including cleft palate when disrupted (Caruana and Bernstein 2001). The haplotype spans nearly the whole *DLG1* gene providing additional evidence for *DLG1* as candidate gene (Figure 2). Of note, strong support for our finding comes from a recent GWAS replication study for the other frequent

cleft phenotype, nsCL/P (Mostowska et al., 2017), which identified a genome-wide significant signal at the *DLG1* locus. In their study the strongest associated SNP ($P = 1.77x10^{-8}$) at the locus is rs338222, which is in LD with our candidate SNP rs395572 ($r^2 > 0.8$). Our candidate SNP rs395572 reached a P-value of $4.54x10^{-7}$ in their study However, the risk alleles at rs338222 and rs395572 reported by Mostowska et al. (2017) differ from the risk alleles in the present nsCPO data, indicating that one allele confers risk for nsCL/P while the other allele confers risk for nsCPO. A similar finding has been recently reported in a population-based study of orofacial clefting (Moreno Uribe et al., 2017). Moreno Uribe et al. (2017) found that the variant rs227731 at the *NOG1* locus showed an association with the opposite risk allele for nsCL/P compared to nsCPO. The functional significance of this observation is unclear. One plausible hypothesis is that the binding of transcription factors to these regions is allele- and context-specific.

Our results support the concept of differences in the genetic background of nsCPO and nsCL/P. First, the 15 novel risk loci for nsCL/P (Yu et al., 2017; Leslie et al., 2017) showed no genome-wide association in the present nsCPO GWAS data. Second, our present study identified only suggestive loci, and nsCPO GWASs have identified only one genome-wide significant finding, namely rs41268753 at the *GRHL3* locus (Beaty et al., 2011; Leslie et al., 2016a). In contrast, GWASs of nsCL/P even when investigating smaller sample sizes identified five genome-wide significant risk loci (Birnbaum et al., 2009; Beaty et al., 2010; Mangold et al., 2010), and many loci approaching genome-wide significance which were confirmed in further studies (Ludwig et al., 2012; Beaty et al., 2013; Sun et al., 2015; Leslie et al., 2016b). Sufficient evidence is therefore now available to consider nsCL/P a complex trait. In contrast, the precise etiological nature of nsCPO remains unclear.

The development of nsCPO may involve fewer genetic factors and more exogenic factors than is the case for nsCL/P. Nonetheless, heritability estimates of approximately 90% and a 50-60 fold increase in the recurrence risk for siblings of nsCPO patients compared to the population - data which strongly resemble estimates for nsCL/P - indicate that genetic factors make a strong contribution to the nsCPO phenotype.

One could raise the hypothesis that the genetic variants involved in nsCPO and nsCL/P differ in terms of their frequency spectrum, and that nsCPO risk variants have lower frequencies. Taking this into consideration common nsCPO risk variants may have been missed with the previous GWASs and also with our approach simply due to sample size, i.e. power issues. Of note, the proven risk variant rs41268753 at the *GRHL3* locus has a frequency of 3.2% in the general population (Non-Finnish

Europeans; gnomAD) (Exome Aggregation Consortium, 2016), and thus is among the low-frequency variants in the European population.

Another hypothesis is that in general, rare variants might play a primary role in nsCPO etiology. These rare variants may be located in genes associated with syndromic forms of orofacial clefting, as shown previously by the present authors for *GRHL3* and the Van-der-Woude syndrome (Mangold et al., 2016). Van der Woude syndrome is characterized by orofacial clefting and lip pits. In the mentioned sequencing study by our group members of the $Mangold_{EU}$ nsCPO sample were found to have Van der Woude syndrome without lip pits. Furthermore, an exome sequencing study of nsCPO identified putative causal variants in *GRHL3* and *CREBBP* (Hoebel et al., 2017). *CREBBP* is associated with Rubinstein-Taybi syndrome, which is characterized among others by CPO (Hennekam et al., 1993), and which might represent a further hypomorphic syndrome with reduced penetrance of additional anomalies. The present and other nsCPO samples may include similar cases. To generate more evidence for this hypothesis, further whole-exome or whole-genome sequencing would be a useful tool to detect rare variants in nsCPO patients.

All of our findings require follow-up in order to assess their true contribution to nsCPO risk. This might involve meta-analyses of the present data and other available nsCPO GWAS datasets, or the use of larger samples in order to maximize statistical power. However, recruitment is hampered by the low prevalence of nsCPO (around 1 in 2,400). An alternative strategy would be to maximize statistical power by increasing marker density and data quality. As the newest generation of haplotype reference panels for imputation – such as the Haplotype Reference Consortium (HRC) (McCarthy et al., 2016) – combine sequencing data from multiple samples, genotype imputation is becoming more accurate for the entire MAF spectrum. In the present study, imputation was performed using the 1000 genomes haplotypes (phase 3). Thus imputation accuracy (and marker density) in general could be increased by imputing the dataset using HRC as a reference panel. Furthermore, variants with MAF < 1% could be considered. Nevertheless, the sample size of the imputed GWAS sample ($EU+Asia_{Beaty}$) was limited, and may have been insufficient in terms of the discovery of associations with low frequency variants.

There are several limitations to the study.

First, in our study we may have overlooked true associated variants in the imputed datasets (EU_{Beaty} and $EU+Asia_{Beaty}$) (Ludwig et al., 2017) due to the stringent selection criteria applied (only SNPs with $P < 1x10^{-5}$ were considered for genotyping). An example for this is again rs41268753 at the *GRHL3* locus, the only proven nsCPO risk variant known to date (Leslie et al., 2016a; Mangold et al., 2016).

We checked the P-value of this variant in both imputed datasets. Of note, the P-value of rs4126875 is higher than $1x10^{-5}$ ($P_{EUBeaty} = 1.18x10^{-4}$; $P_{EU+AsiaBeaty} = 1.42x10^{-4}$) (data not shown). This is above our set P-value threshold in the SNP selection procedure and demonstrates that our approach can miss a true finding.

Second, although CPO often occurs as part of a syndrome, it may be misdiagnosed as nonsyndromic when reduced penetrance of additional symptoms occurs. Therefore, the present GWAS and genotyped samples may have included syndromic cases with a monogenetic background. If so, this would have led to a loss of statistical power. However, Moreno Uribe et al. compared locus effects in nonsyndromic (isolated) and syndromic cases, and showed that for most loci, the associations were similar in terms of direction and magnitude (Moreno Uribe et al., 2017). Notably, the Moreno Uribe analysis was performed predominantly in nsCL/P, since only one nsCPO risk locus was analyzed.

Third, the study is based on imputed GWAS data of trios. Therefore not all of the meta-analyses provided effect sizes, which hampers the interpretation of results.

A fourth limitation are the sizes and ethnicities of the samples used in this study. Sample sizes might be too small to detect true associations. Asian trios were part of the dataset that led to candidate SNPs, but none of the genotyped samples was of Asian ethnicity. Variants with a population-specific effect restricted to Asians may have been missed by our approach.

In conclusion, our findings suggest that there is a limited contribution of common variants to nsCPO. Rare variants may play a more substantial role in nsCPO etiology than in nsCL/P. Whole-exome/genome sequencing studies are necessary to identify such rare causative variants. However, our study identified two interesting suggestive nsCPO risk variants. The first finding rs395572 is located at a recently identified risk locus for nsCL/P at 3q29 and lies in high LD with the genome-wide significant top SNP at the locus (Mostowska et al., 2017). Although it has to be noted that the SNP shows the opposite risk allele in our study. The other finding rs6809420 at 3q13 is located in a potential p300 binding site, a coactivator that is known to be involved in craniofacial development. Follow-up studies with increased sample sizes could further elucidate whether these variants are truly associated with nsCPO.

Acknowledgements:

The authors thank the patients and their families, and all control individuals, for their participation in the study. We acknowledge the valuable assistance of all clinical and laboratory staff. We thank the

German support group for individuals with cleft lip and/or palate (Wolfgang Rosenthal Gesellschaft), and the Network for Orofacial Clefts Research, Prevention, and Treatment (EUROCleftNet) for assistance with recruitment. Furthermore, the authors would like to thank the Genome Aggregation Database (gnomAD) and the groups that provided exome and genome variant data to this resource. A full list of contributing groups can be found at http://gnomad.broadinstitute.org/about. This work was supported by the European Science Foundation.

Conflict of interest:

The authors have no conflict of interest.

References:

Aldhorae KA, Böhmer AC, Ludwig KU, Esmail AHA, Al-Hebshi NN, Lippke B, Gölz L, Nöthen MM, Daratsianos N, Knapp M, Jäger A, Mangold E. 2014. Nonsyndromic cleft lip with or without cleft palate in Arab populations: Genetic analysis of 15 risk loci in a novel case-control sample recruited in Yemen. Birth Defects Res Part A - Clin Mol Teratol 100:307–313.

Beaty TH, Murray JC, Marazita ML, Munger RG, Ruczinski I, Hetmanski JB, Liang KY, Wu T, Murray T, Fallin MD, Redett RA, Raymond G, Schwender H, Jin S, Cooper ME, Dunnwald M, Mansilla MA, Leslie E, Bullard S, Lidral AC, Moreno LM, Menezes R, Vieira AR, Petrin A, Wilcox AJ, Lie RT, Jabs EW, Wu-Chou YH, Chen PK, Wang H, Ye X, Huang S, Yeow V, Chong SS, Jee SH, Shi B, Christensen K, Melbye M, Doheny KF, Pugh EW, Ling H, Castilla EE, Czeizel AE, Ma L, Field LL, Brody L, Pangilinan F, Mills JL, Molloy AM, Kirke PN, Scott JM, Arcos-Burgos M, Scott AF. 2010. A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. Nat Genet 42:525–9.

Beaty TH, Ruczinski I, Murray JC, Marazita ML, Munger RG, Hetmanski JB, Murray T, Redett RJ, Fallin MD, Liang KY, Wu T, Patel PJ, Jin S, Zhang TX, Schwender H, Wu-Chou YH, Chen PK, Chong SS, Cheah F, Yeow V, Ye X, Wang H, Huang S, Jabs EW, Shi B, Wilcox AJ, Lie RT, Jee SH, Christensen K, Doheny KF, Pugh EW, Ling H, Scott AF. 2011. Evidence for gene-environment interaction in a genome wide study of nonsyndromic cleft palate. Genet Epidemiol 35:469–478.

Beaty TH, Taub MA, Scott AF, Murray JC, Marazita ML, Schwender H, Parker MM, Hetmanski JB, Balakrishnan P, Mansilla MA, Mangold E, Ludwig KU, Nöthen MM, Rubini M, Elcioglu N, Ruczinski I. 2013. Confirming genes influencing risk to cleft lip with/without cleft palate in a case-parent trio study. Hum Genet 132:771-781.

Becker T, Knapp M. 2004. A powerful strategy to account for multiple testing in the context of haplotype analysis. Am J Hum Genet 75:561–570.

Bhattacherjee V, Horn KH, Singh S, Webb CL, Pisano MM, Greene RM. 2009. CBP/p300 and associated transcriptional co-activators exhibit distinct expression patterns during murine craniofacial and neural tube development. Int J Dev Biol 53:1097–1104.

Birnbaum S, Ludwig KU, Reutter H, Herms S, Steffens M, Rubini M, Baluardo C, Ferrian M, Almeida de Assis N, Alblas MA, Barth S, Freudenberg J, Lauster C, Schmidt G, Scheer M, Braumann B, Berge SJ, Reich RH, Schiefke F, Hemprich A, Potzsch S, Steegers-Theunissen RP, Potzsch B, Moebus S, Horsthemke B, Kramer FJ, Wienker TF, Mossey PA, Propping P, Cichon S, Hoffmann P, Knapp M, Nöthen MM, Mangold E. 2009. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. Nat Genet 41:473-477.

Cobat, A, Abel L., Alcaïs A, Schurr E. 2014. A general efficient and flexible approach for genome-wide association analyses of imputed genotypes in family-based designs. Genet Epidemiol 38:560-571.

Ding H-L, Hooper JE, Batzel P, Eames BF, Postlethwait JH, Artinger KB, Clouthier DE. 2016. MicroRNA Profiling during Craniofacial Development: Potential Roles for Mir23b and Mir133b. Front Physiol 7:281.

Exome Aggregation Consortium. 2016. Nature 536:285-291.

Grosen D, Bille C, Petersen I, Skytthe A, Bornemann J Von, Pedersen JK, Murray JC, Christensen K. 2011. Risk of Oral Clefts in twins. Epidemiology 22:313–319.

Hennekam RC, Tilanus M, Hamel BC, Voshart-van Heeren H, Mariman EC, van Beersum SE, van den Boogaard MJ, Breuning MH. 1993. Deletion at chromosome 16p13.3 as a cause of Rubinstein-Taybi syndrome: clinical aspects. Am J Hum Genet 52:255–262.

Hoebel AK, Drichel D, van de Vorst M, Böhmer AC, Sivalingam S, Ishorst N, Klamt J, Gölz L, Alblas M, Maaser A, Keppler K, Zink AM, Dixon MJ, Dixon M, Hemprich A, Kruse T, Graf I, Dunsche A, Schmidt G, Daratsianos N, Nowak S, Aldhorae KA, Nöthen MM, Knapp M, Thiele H, Gilissen C, Reutter H, Hoischen A, Mangold E, Ludwig KU. 2017. Candidate Genes for Nonsyndromic Cleft Palate Detected by Exome Sequencing. J Dent Res 96:1314-1321.

Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. 2008. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24:2938–2939.

Kheradpour P, Kellis M. 2013. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. Nucleic Acids Res 42:2976–2987.

Leslie EJ, Liu H, Carlson JC, Shaffer JR, Feingold E, Wehby G, Laurie CA, Jain D, Laurie CC, Doheny KF, McHenry T, Resick J, Sanchez C, Jacobs J, Emanuele B, Vieira AR, Neiswanger K, Standley J, Czeizel AE, Deleyiannis F, Christensen K, Munger RG, Lie RT, Wilcox A, Romitti PA, Field LL, Padilla CD, Cutiongco-de la Paz EM, Lidral AC, Valencia-Ramirez LC, Lopez-Palacio AM, Valencia DR, Arcos- Burgos M, Castilla EE, Mereb JC, Poleta FA, Orioli IM, Carvalho FM, Hecht JT, Blanton SH, Buxó CJ, Butali A, Mossey PA, Adeyemo WL, James O, Braimah RO, Aregbesola BS, Eshete MA, Deribew M, Koruyucu M, Seymen F, Ma L, de Salamanca JE, Weinberg SM, Moreno L, Cornell RA, Murray JC, Marazita ML. 2016a. A Genome-wide Association Study of Nonsyndromic Cleft Palate Identifies an Etiologic Missense Variant in GRHL3. Am J Hum Genet 98:744–754.

Leslie EJ, Carlson JC, Shaffer JR, Feingold E, Wehby G, Laurie CA, Jain D, Laurie CC, Doheny KF, McHenry T, Resick J, Sanchez C, Jacobs J, Emanuele B, Vieira AR, Neiswanger K, Lidral AC, Valencia-Ramirez LC, Lopez-Palacio AM, Valencia DR, Arcos-Burgos M, Czeizel AE, Field LL, Padilla CD, Cutiongco-de la Paz EM, Deleyiannis F, Christensen K, Munger RG, Lie RT, Wilcox A, Romitti PA, Castilla EE, Mereb JC, Poletta FA, Orioli IM, Carvalho FM, Hecht JT, Blanton SH, Buxo CJ, Butali A, Mossey PA, Adeyemo WL, James O, Braimah RO, Aregbesola BS, Eshete MA, Abate F, Koruyucu M, Seymen F, Ma L, de Salamanca JE, Weinberg SM, Moreno L, Murray JC, Marazita ML. 2016b. A multi-ethnic genome-wide association study identifies novel loci for non-syndromic cleft lip with or without cleft palate on 2p24.2, 17q23 and 19q13. Hum Mol Genet 25:2862-2872.

Leslie EJ, Carlson JC, Shaffer JR, Butali A, Buxo CJ, Castilla EE, Christensen K, Deleyiannis FW, Leigh Field L, Hecht JT, Moreno L, Orioli IM, Padilla C, Vieira AR, Wehby GL, Feingold E, Weinberg SM, Murray JC, Beaty TH, Marazita ML. 2017. Genome-wide meta-analyses of nonsyndromic orofacial clefts identify novel associations between FOXE1 and all orofacial clefts, and TP63 and cleft lip with or without cleft palate. Hum Genet 136:275-286.

Li D, Zhang H, Ma L, Han Y, Xu M, Wang Z, Jiang H, Zhang W, Wang L, Pan Y. 2016. Associations between microRNA binding site SNPs in FGFs and FGFRs and the risk of non-syndromic orofacial cleft. Sci Rep 6:31054.

Ludwig KU, Ahmed ST, Böhmer AC, Sangani NB, Varghese S, Klamt J, Schuenke H, Gultepe P, Hofmann A, Rubini M, Aldhorae KA, Steegers-Theunissen RP, Rojas-Martinez A, Reiter R, Borck G, Knapp M, Nakatomi M, Graf D, Mangold E, Peters H. 2016. Meta-analysis Reveals Genome-Wide Significance at 15q13 for Nonsyndromic Clefting of Both the Lip and the Palate, and Functional Analyses Implicate GREM1 As a Plausible Causative Gene. PLoS genetics 12:e1005914.

Ludwig KU, Böhmer AC, Rubini M, Mossey PA, Herms S, Nowak S, Reutter H, Alblas MA, Lippke B, Barth S, Paredes-Zenteno M, Munoz-Jimenez SG, Ortiz-Lopez R, Kreusch T, Hemprich A, Martini M, Braumann B, Jager A, Potzsch B, Molloy A, Peterlin B, Hoffmann P, Nöthen MM, Rojas-Martinez A, Knapp M, Steegers-Theunissen RP, Mangold E. 2014. Strong association of variants around FOXE1 and orofacial clefting. J Dent Res 93:376-381.

Ludwig KU, Böhmer AC, Bowes J, Nikolic M, Ishorst N, Wyatt N, Hammond NL, Golz L, Thieme F, Barth S, Schuenke H, Klamt J, Spielmann M, Aldhorae K, Rojas-Martinez A, Nöthen MM, Rada-Iglesias A, Dixon MJ, Knapp M, Mangold E. 2017. Imputation of Orofacial Clefting Data Identifies Novel Risk Loci and Sheds Light on the Genetic Background of Cleft Lip +/- Cleft Palate and Cleft Palate Only. Hum Mol Genet 26:829-842.

Ludwig KU, Mangold E, Herms S, Nowak S, Reutter H, Paul A, Becker J, Herberz R, AlChawa T, Nasser E, Böhmer AC, Mattheisen M, Alblas MA, Barth S, Kluck N, Lauster C, Braumann B, Reich RH, Hemprich A, Potzsch S, Blaumeiser B, Daratsianos N, Kreusch T, Murray JC, Marazita ML, Ruczinski I, Scott AF, Beaty TH, Kramer FJ, Wienker TF, Steegers-Theunissen RP, Rubini M, Mossey PA, Hoffmann P, Lange C, Cichon S, Propping P, Knapp M, Nöthen MM. 2012. Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. Nat Genet 44:968-971.

Mangold E, Böhmer AC, Ishorst N, Hoebel AK, Gültepe P, Schuenke H, Klamt J, Hofmann A, Gölz L, Raff R, Tessmann P, Nowak S, Reutter H, Hemprich A, Kreusch T, Kramer FJ, Braumann B, Reich R, Schmidt G, Jäger A, Reiter R, Brosch S, Stavusis J, Ishida M, Seselgyte R, Moore GE, Nöthen MM, Borck G, Aldhorae KA, Lace B, Stanier P, Knapp M, Ludwig KU. 2016. Sequencing the GRHL3 Coding Region Reveals Rare Truncating Mutations and a Common Susceptibility Variant for Nonsyndromic Cleft Palate. Am J Hum Genet 98:755–762.

Mangold E, Ludwig KU, Birnbaum S, Baluardo C, Ferrian M, Herms S, Reutter H, de Assis NA, Chawa TA, Mattheisen M, Steffens M, Barth S, Kluck N, Paul A, Becker J, Lauster C, Schmidt G, Braumann B, Scheer M, Reich RH, Hemprich A, Pötzsch S, Blaumeiser B, Moebus S, Krawczak M, Schreiber S, Meitinger T, Wichmann HE, Steegers-Theunissen RP, Kramer FJ, Cichon S, Propping P, Wienker TF, Knapp M, Rubini M, Mossey PA, Hoffmann P, Nöthen MM. 2010. Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. Nat. Genet. 42:24–26.

Mangold E, Ludwig KU, Nöthen MM. 2011. Breakthroughs in the genetics of orofacial clefting. Trends Mol Med 17:725-733.

McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, Luo Y, Sidore C, Kwong A, Timpson N, Koskinen S, Vrieze S, Scott LJ, Zhang H, Mahajan A, Veldink J, Peters U, Pato C, van Duijn CM, Gillies CE, Gandin I, Mezzavilla M, Gilly A, Cocca M, Traglia M, Angius A, Barrett JC, Boomsma D, Branham K, Breen G, Brummett CM, Busonero F, Campbell H, Chan A, Chen S, Chew E, Collins FS, Corbin LJ, Smith GD, Dedoussis G, Dorr M, Farmaki AE, Ferrucci L, Forer L, Fraser RM, Gabriel S, Levy S, Groop L, Harrison T, Hattersley A, Holmen OL, Hveem K, Kretzler M, Lee JC, McGue M, Meitinger T, Melzer D, Min JL, Mohlke KL, Vincent JB, Nauck M, Nickerson D, Palotie A, Pato M, Pirastu N, McInnis M, Richards JB, Sala C, Salomaa V, Schlessinger D, Schoenherr S, Slagboom PE, Small K, Spector T, Stambolian D, Tuke M, Tuomilehto J, Van den Berg

LH, Van Rheenen W, Volker U, Wijmenga C, Toniolo D, Zeggini E, Gasparini P, Sampson MG, Wilson JF, Frayling T, de Bakker PI, Swertz MA, McCarroll S, Kooperberg C, Dekker A, Altshuler D, Willer C, Iacono W, Ripatti S, Soranzo N, Walter K, Swaroop A, Cucca F, Anderson CA, Myers RM, Boehnke M, McCarthy MI, Durbin R; Haplotype Reference Consortium. 2016. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 48:1279–1283.

Moreno LM, Mansilla MA, Bullard SA, Cooper ME, Busch TD, Machida J, Johnson MK, Brauer D, Krahn K, Daack-Hirsch S, L'Heureux J, Valencia-Ramirez C, Rivera D, Lopez AM, Moreno MA, Hing A, Lammer EJ, Jones M, Christensen K, Lie RT, Jugessur A, Wilcox AJ, Chines P, Pugh E, Doheny K, Arcos-Burgos M, Marazita ML, Murray JC, Lidral AC. 2009. FOXE1 association with both isolated cleft lip with or without cleft palate, and isolated cleft palate. Hum Mol Genet 18:4879-4896.

Moreno Uribe LM, Fomina T, Munger RG, Romitti PA, Jenkins MM, Gjessing HK, Gjerdevik M, Christensen K, Wilcox AJ, Murray JC, Lie RT, Wehby GL. 2017. A Population-Based Study of Effects of Genetic Loci on Orofacial Clefts. J Dent Res 96:1322-1329.

Mossey PA, Little J, Steegers-Theunissen R, Molloy A, Peterlin B, Shaw WC, Johnson C, FitzPatrick DR, Franceschelli P, Rubini M. 2017. Genetic Interactions in Nonsyndromic Orofacial Clefts in Europe—EUROCRAN Study. Cleft Palate Craniofac J 54:623-630.

Mostowska A, Gaczkowska A, Żukowskib K, Ludwig KU, Hozyasze KK, Wójcickif P, Mangold E, Böhmer AC, Heilmann-Heimbach S, Knapp M, Zadurskai M, Biedziakj B, Budner M, Lasota A, Daktera-Micker A, Jagodziński PP. 2017. Common variants in DLG1 locus are associated with non-syndromic cleft lip with or without cleft palate. Clin Genet doi: 10.1111/cge.13141.

Rahimov F, Marazita ML, Visel A, Cooper ME, Hitchler MJ, Rubini M, Domann FE, Govil M, Christensen K, Bille C, Melbye M, Jugessur A, Lie RT, Wilcox AJ, Fitzpatrick DR, Green ED, Mossey PA, Little J, Steegers-Theunissen RP, Pennacchio LA, Schutte BC, Murray JC. 2008. Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. Nat Genet 40:1341-1347.

Sun Y, Huang Y, Yin A, Pan Y, Wang Y, Wang C, Du Y, Wang M, Lan F, Hu Z, Wang G, Jiang M, Ma J, Zhang X, Ma H, Ma J, Zhang W, Huang Q, Zhou Z, Ma L, Li Y, Jiang H, Xie L, Jiang Y, Shi B, Cheng J, Shen H, Wang L, Yang Y. 2015. Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. Nat Commun 6:6414.

The 1000 genomes Project Consortium. 2015. A global reference for human genetic variation. Nature 526:68-74.

Thieme F, Ludwig KU. 2017. The Role of Noncoding Genetic Variation in Isolated Orofacial Clefts. J Dent Res 96:1238-1247.

Wang J, Bai Y, Li H, Greene SB, Klysik E, Yu W, Schwartz RJ, Williams TJ, Martin JF. 2013. MicroRNA-17-92, a direct Ap-2α transcriptional target, modulates T-box Factor activity in orofacial clefting. PLoS Genet 9:e1003785.

Willer CJ, Li Y, Abecasis GR. 2010. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 26:2190-1.

Yu Y, Zuo X, He M, Gao J, Fu Y, Qin C, Meng L, Wang W, Song Y, Cheng Y, Zhou F, Chen G, Zheng X, Wang X, Liang B, Zhu Z, Fu X, Sheng Y, Hao J, Liu Z, Yan H, Mangold E, Ruczinski I, Liu J, Marazita ML, Ludwig KU, Beaty TH, Zhang X, Sun L, Bian Z. 2017. Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. Nat Commun 8:14364.

Table and Figure legends:

Figure 1: Workflow of the present study.

Ishorst et al., Candidate loci association study in nsCPO

Figure 2: Regional association plot of rs395572 locus +/- 500kb. $P_{EUBeaty}$ = P-value from imputed GWAS of European trios. P_{MetaEU} = P-value from imputed GWAS of European trios and all genotyped samples with European ancestry.

Figure 3: Regional association plot of rs6809420 locus +/- 500kb. $P_{EU+AsiaBeaty}$ = P-value from imputed GWAS of all trios. $P_{MetaAll}$ = P-value from imputed GWAS of all trios and all genotyped samples. Genes *TUSC1* and *IGSF11* are located 1 Mb upstream and 1.2 Mb downstream of rs6809420, respectively. Rs7633157 belongs to another locus indicated by LD (represented by dashed line) and failed genotyping.

Table 1: Association results for SNPs with nominally significant associations in at least one of three genotyped samples and/or their meta-analysis ($Meta_{geno}$).

Table 2: Results of European meta-analysis (SNPs with a decreased P-value in meta-analysis as compared to intitial imputed GWAS dataset (EU_{Beaty}) are shown.

Table 3: Results of meta-analysis of all genotyped samples and $EU+Asia_{Beaty}$ (SNPs with a decreased P-value after meta-analysis in comparison to the initial imputed GWAS dataset ($EU+Asia_{Beaty}$) are shown here.)

Supplementary Figure S1: Regional association plot of rs595533 locus +/- 500kb. $P_{EU+AsiaBeaty}$ = P-value from imputed GWAS of all trios. $P_{MetaAll}$ = P-value from imputed GWAS of all trios and all genotyped samples.

Supplementary Table S1: Gender distribution and origin of the three genotyped samples

Supplementary Table S2: 83 SNPs with $P<1x10^{-5}$ in imputed datasets EU_{Beaty} and $EU+Asia_{Beaty}$

Supplementary Table S3: Genotyped SNPs from imputed GWAS (EU_{Beaty}, $EU+Asia_{Beaty}$)

Supplementary Table S4: SNPs from Leslie et al. (2016a)

Supplementary Table S5: SNPs finally selected for genotyping

Supplementary Table S6: Relative risks of genotyped samples and meta-analysis of all genotyped samples ($Meta_{geno}$) derived from a fixed-effects based meta-analysis.

Supplementary Table S7: Association P-values of EU_{Beaty} and $EU+Asia_{Beaty}$ at the 14 novel nsCL/P loci from Yu et al. (2017) and one novel locus from Leslie et al. (2017) at lead SNP level

Supplementary Table S8: Association results for two European ($EU_{Mangold}$ and EU_{Mossey}) and one Arabian ($Arab_{Aldhorae}$) nsCPO sample and the corresponding meta-analysis

Supplementary Table S9: Association results for the European nsCPO samples ($EU_{Mangold}$ and EU_{Mossey}), imputation of the European nsCPO trios (EU_{Beaty}) and the corresponding meta-analysis ($Meta_{EU}$)

Supplementary Table S10: Association results for the European and the Arabian nsCPO samples ($EU_{Mangold,}$ EU_{Mossey} and $Arab_{Aldhorae}$), imputation of the nsCPO trios ($EU+Asia_{Beaty}$) and the corresponding meta-analysis ($Meta_{All}$).

Supplementary Table S11: Association P-values of EU_{Beaty} at 14 novel nsCL/P loci from Yu et al. (2017) and one novel locus from Leslie et al. (2017). For each locus all SNPs with P-value < $1x10^{-3}$ (info-score > 0.8) or if none present the SNPs with the two lowest P-values (info-score > 0.8) are shown.

Supplementary Table S12: Association P-values of $EU+Asia_{Beaty}$ at 14 novel nsCL/P loci from Yu et al. (2017) and one novel locus from Leslie et al. (2017). For each locus all SNPs with P-value < $1x10^{-3}$ (info-score > 0.8) or if none present the SNPs with the two lowest P-values (info-score > 0.8) are shown.

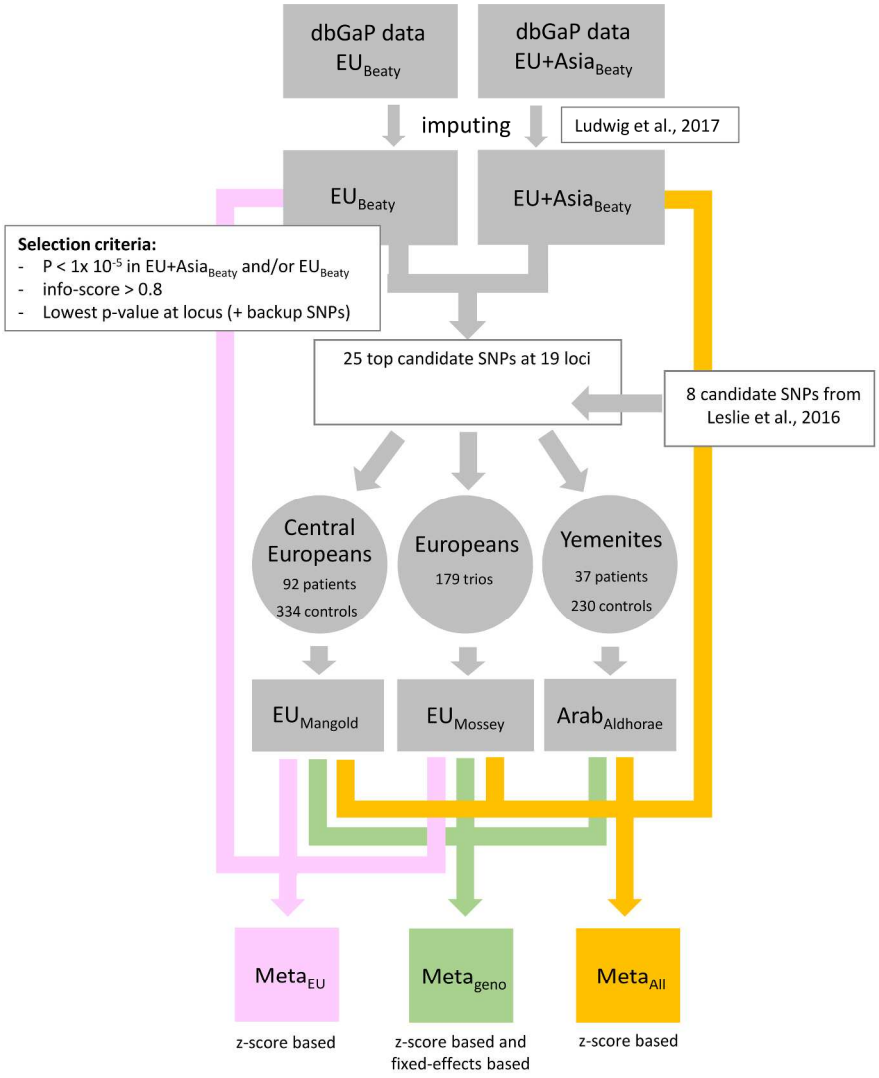


Figure 1: Workflow of the present study.

233x318mm (300 x 300 DPI)

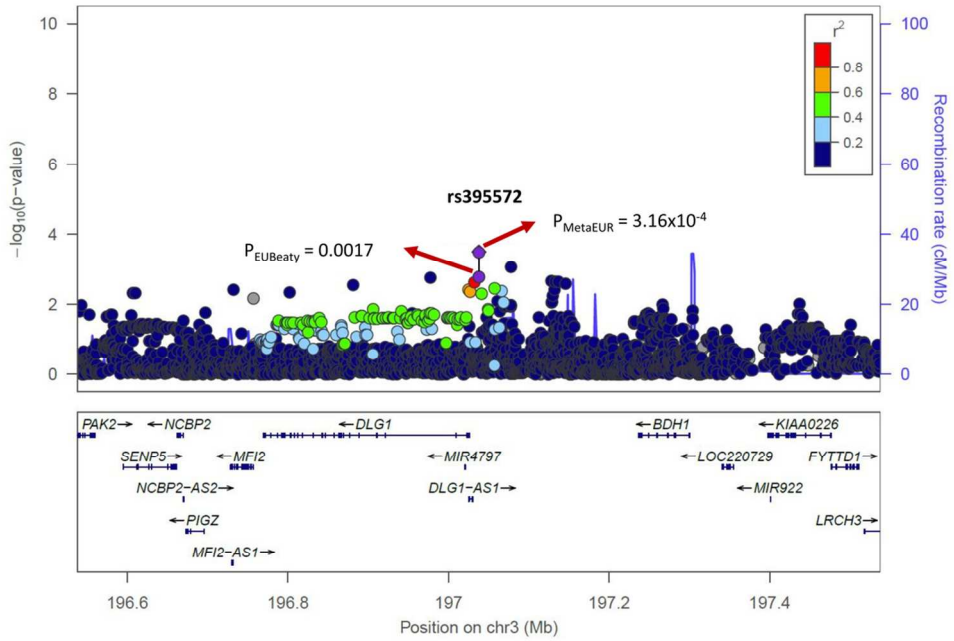


Figure 2: Regional association plot of rs395572 locus +/- 500kb. $P_{EUBeaty}$ = P-value from imputed GWAS of European trios. P_{MetaEU} = P-value from imputed GWAS of European trios and all genotyped samples with European ancestry.

109x74mm (300 x 300 DPI)

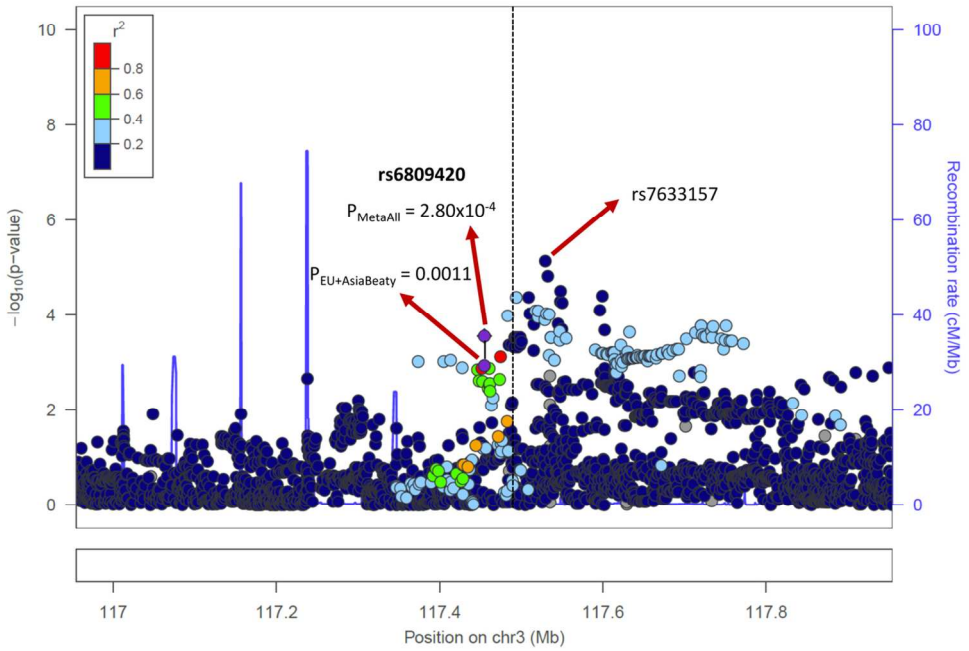


Figure 3: Regional association plot of rs6809420 locus +/- 500kb. $P_{EU+AsiaBeaty}$ = P-value from imputed GWAS of all trios. $P_{MetaAll}$ = P-value from imputed GWAS of all trios and all genotyped samples. Genes *TUSC1* and *IGSF11* are located 1 Mb upstream and 1.2 Mb downstream of rs6809420, respectively. Rs7633157 belongs to another locus indicated by LD (represented by dashed line) and failed genotyping.

119x91mm (300 x 300 DPI)

Table 1 Association results for SNPs with nominally significant associations in at least one of three genotyped samples and/or their meta-analysis ($Meta_{geno}$)

SNP information				Association P-values[c]			
SNP	Chr	Pos (hg19)	Risk/other allele[a]	$EU_{Mangold}$ P_{Trend}	$Arab_{Aldhorae}$ P_{Trend}	EU_{Mossey} P_{TDT}	$Meta_{geno}$
rs6747560	2	32784766	G/T[b]	**0.007**[d]	*0.414*	0.320	**0.046**[d]
rs1283107	3	106961320	G/C[b]	0.383	0.281	0.142	**0.049**[d]
rs6809420	3	117454822	T/G	0.320	**0.024**	0.647	0.066
rs6805813	3	187595294	G/A[b]	0.733	**0.019**[d]	*0.535*	0.456
rs395572	3	197037952	A/G	0.309	***0.042***[e]	0.058	0.278

SNP = single nucleotide polymorphism; Chr = Chromosome; Pos (hg19) = postion in human reference genome version 19

Nominally significant P-values are given in bold; P-values given in italics indicate an opposite effect direction as in the meta-analysis ($Meta_{geno}$)

[a] risk allele according to $Meta_{geno}$

[b] different risk allele as compared to imputed nsCPO datasets (EU_{Beaty} and $EU+Asia_{Beaty}$)

[c] Association P-values for $EU_{Mangold}$ and $Arab_{Aldhorae}$ cohorts were calculated using the Armitage trend test (P_{trend}); Association P-values for EU_{Mossey} were calculated using the transmission disequilibrium test (P_{TDT})

[d] SNP shows nominally significant association but in the opposite effect direction to that observed in the initial GWAS data

[e] SNP shows nominally significant association but in an opposite effect direction as the initial GWAS data and $Meta_{geno}$

Table 2 Results of European meta-analysis (SNPs with a decreased P-value in meta-analysis as compared to intitial imputed GWAS dataset (EU_{Beaty}) are shown.)

SNP information				Association P-values[c]			
SNP	Chr	Pos (hg19)	Risk/other allele[b]	$EU_{Mangold}$ P_{Trend}	EU_{Mossey} P_{TDT}	EU_{Beaty}	$Meta_{EU}$
rs395572	3	197037952	A/G	0.309	0.058	0.002	**$3.16x10^{-4}$**
rs7732608[a]	5	155799466	A/G	0.588	0.117	0.496	0.106
rs3740617[a]	11	33881016	C/T	0.062	0.651	0.417	0.081

SNP = single nucleotide polymorphism; Chr = Chromosome; Pos (hg19) = postion in human reference genome version 19

Nominally significant P-values of the meta-analysis are given in bold

[a] SNP derived from Leslie et al. 2016

[b] risk allele according to $Meta_{EU}$

[c] Association P-values for $EU_{Mangold}$ and EU_{Beaty} were calculated using the Armitage trend test (P_{trend}); Association p-values for $EU_{Rubini\&Mossey}$ were calculated using the transmission disequilibrium test (P_{TDT})

Table 3 Results of meta-analysis of all genotyped samples and EU+Asia$_{Beaty}$ (SNPs with a decreased P-value after meta-analysis in comparison to the initial imputed GWAS dataset (EU+Asia$_{Beaty}$) are shown here.

SNP information				Association P-values[c]				
SNP	Chr	Pos (hg19)	Risk/other allele[b]	EU$_{Mangold}$ P$_{Trend}$	Arab$_{Aldhorae}$ P$_{Trend}$	EU$_{Mossey}$ P$_{TDT}$	EU+Asia$_{Beaty}$	Meta$_{All}$
rs6809420	3	117454822	T/G	0.320	0.024	0.647	0.001	**2.80x10^{-4}**
rs7732608[a]	5	155799466	A/G	0.588	*0.434*	0.117	0.498	0.218
rs9347594[a]	6	162587892	C/T	0.084	0.789	0.824	0.728	0.249
rs3740617[a]	11	33881016	C/T	0.062	*0.626*	0.651	0.640	0.232
rs595533[a]	18	66441589	T/C	0.147	0.961	0.712	0.087	**0.041**

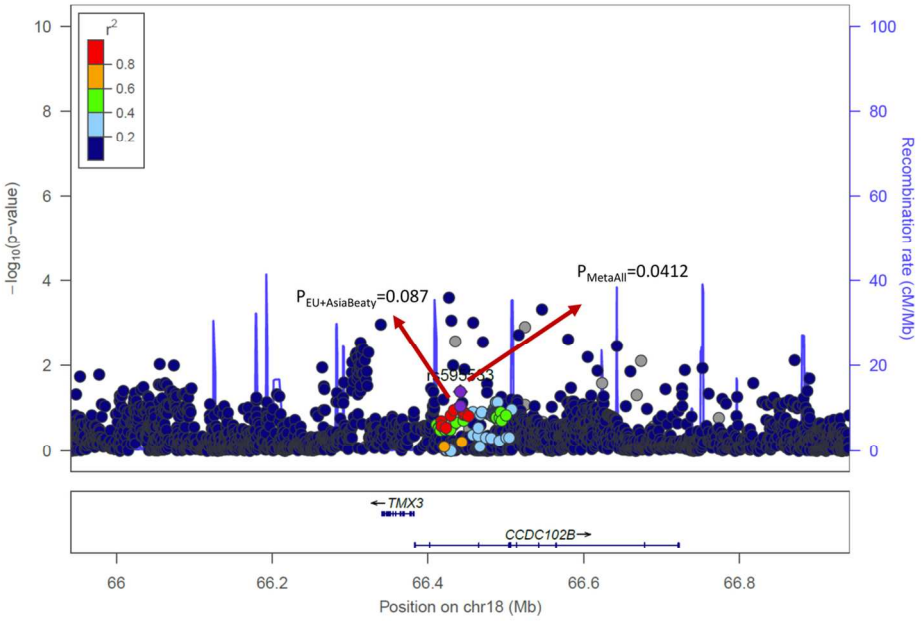
SNP = single nucleotide polymorphism; Chr = Chromosome; Pos (hg19) = postion in human reference genome version 19

Nominally significant P-values of the meta-analysis are given in bold; P-values given in italics indicate an opposite effect direction as in the meta-analysis (Meta$_{All}$)

[a] SNP derived from Leslie et al. 2016

[b] risk allele according to Meta$_{All}$

[c] Association P-values for EU$_{Mangold}$, Arab$_{Aldhorae}$ and EU$_{Beaty}$ were calcualted using the Armitage trend test (P$_{Trend}$); Association P-values for EU$_{Mossey}$ were calcualted using the transmission disequilibrium test (P$_{TDT}$)



Supplementary Figure S1: Regional association plot of rs595533 locus +/- 500kb. $P_{EU+AsiaBeaty}$ = P-value from imputed GWAS of all trios.

119x85mm (300 x 300 DPI)